Ch2 – L5.1

ATP chromatin remodelers

Cytosine methylation analysis genome-wide

Chromatin states are quite dynamic, also at heterochromatin

Example: in heterochromatic domains, we always find HDACs and HMTs that keep the repressed status against the action of HATs and HKDMs

It is known for example that if you treat cells with the histone deacetylase inhibitor «valproic acid», a number of heterochromatic domains loosen and several genes are re-activated, leading to partial de-differentiation.

Another concept that is linked to «chromatin dynamics» is that:

during the establishment of chromatin domains
or gene activation
or transcription
or DNA replication
or preparation to mitosis

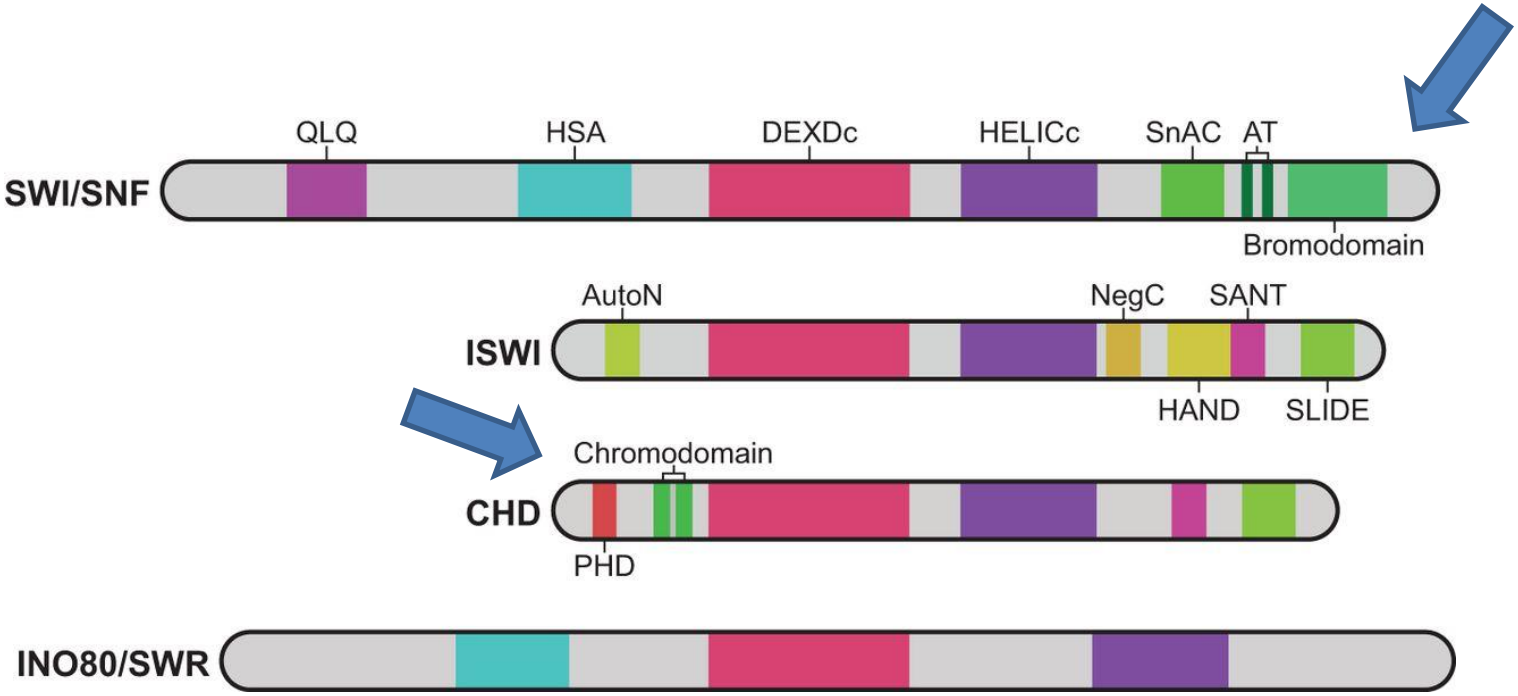
nucleosomes must be «moved» or re-ordered

We call this «chromatin remodelling»

There are a number of ATP-dependent enzymes that execute this nucleosomal movements

ATP-dependent crhomatin remodelers

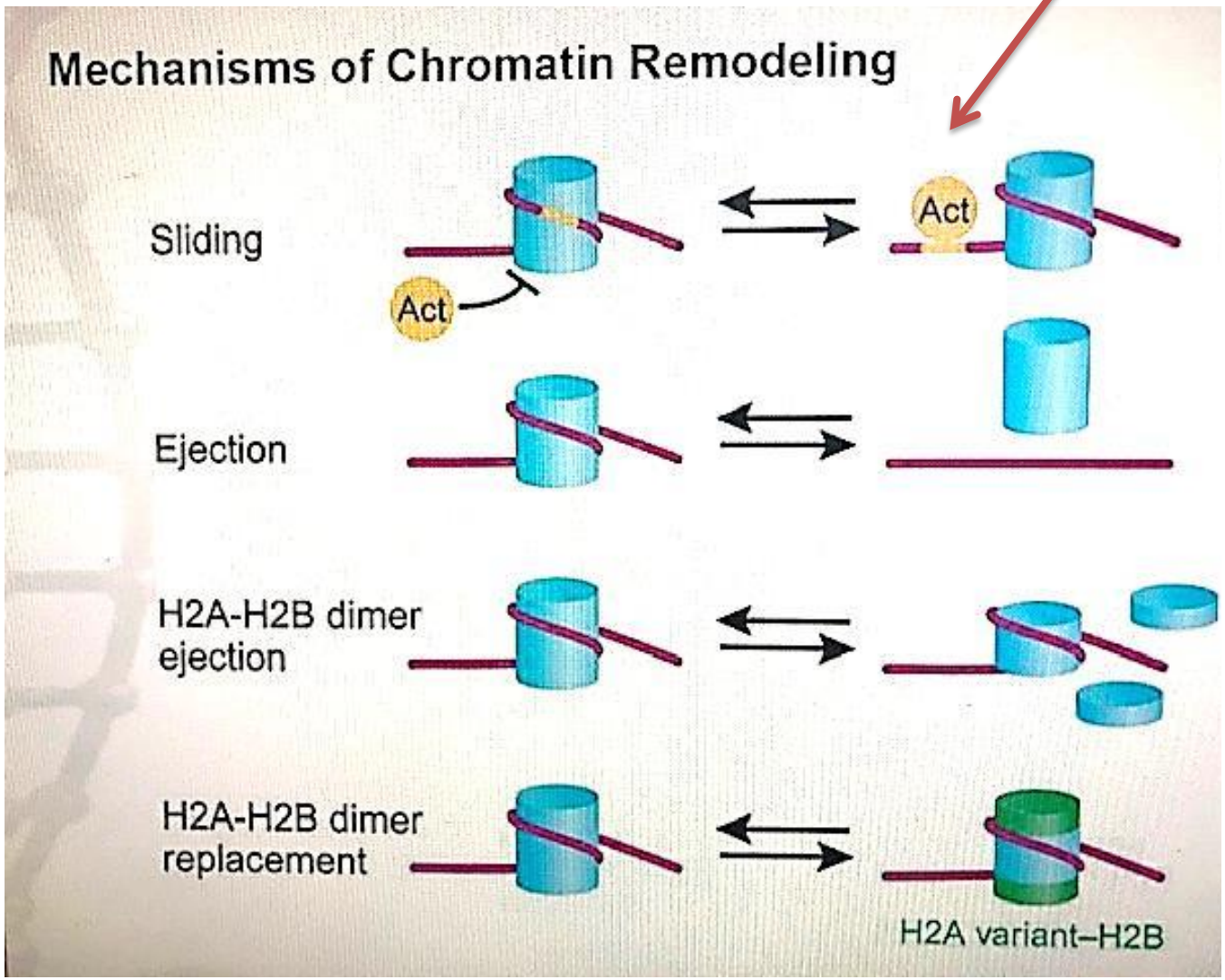
Families of ATP-dependent chromatin remodelers



The domain structures of SNF2 family proteins. The domain organization of the catalytic subunits of SWI/SNF, ISWI, CHD and INO80/SWR subfamilies of chromatin remodelers are shown. All of these subunits are SNF2 family proteins. They all contain an **ATPase domain**, which consists of **DEXDc and HELICc** domains, with each subfamily possessing additional domains. SWI/SNF proteins, for example, are defined by the presence of an N-terminal helicase-SANT (HSA) domain and a C-terminal **bromodomain**. These proteins also contain QLQ and SNF2 ATP-coupling (SnAC) domains, as well as two A-T hook motifs. By contrast, ISWI proteins harbor a C-terminal SANT domain as well as SANT-like ISWI (SLIDE) and HAND domains. They also contain AutoN and NegC regulatory domains. CHD proteins are defined by the presence of tandem N-terminal **chromodomains**, with some family members containing N-terminal plant homeodomain (PHD) domains. INO80R/SWR proteins notably contain a split ATPase domain, with a spacer between the DEXDc and HELICc domains.

From: Hota & Bruneau Development 2016 143: 2882-2897; doi: 10.1242/dev.128892

This is the main test in vitro: accessibility of a TF.

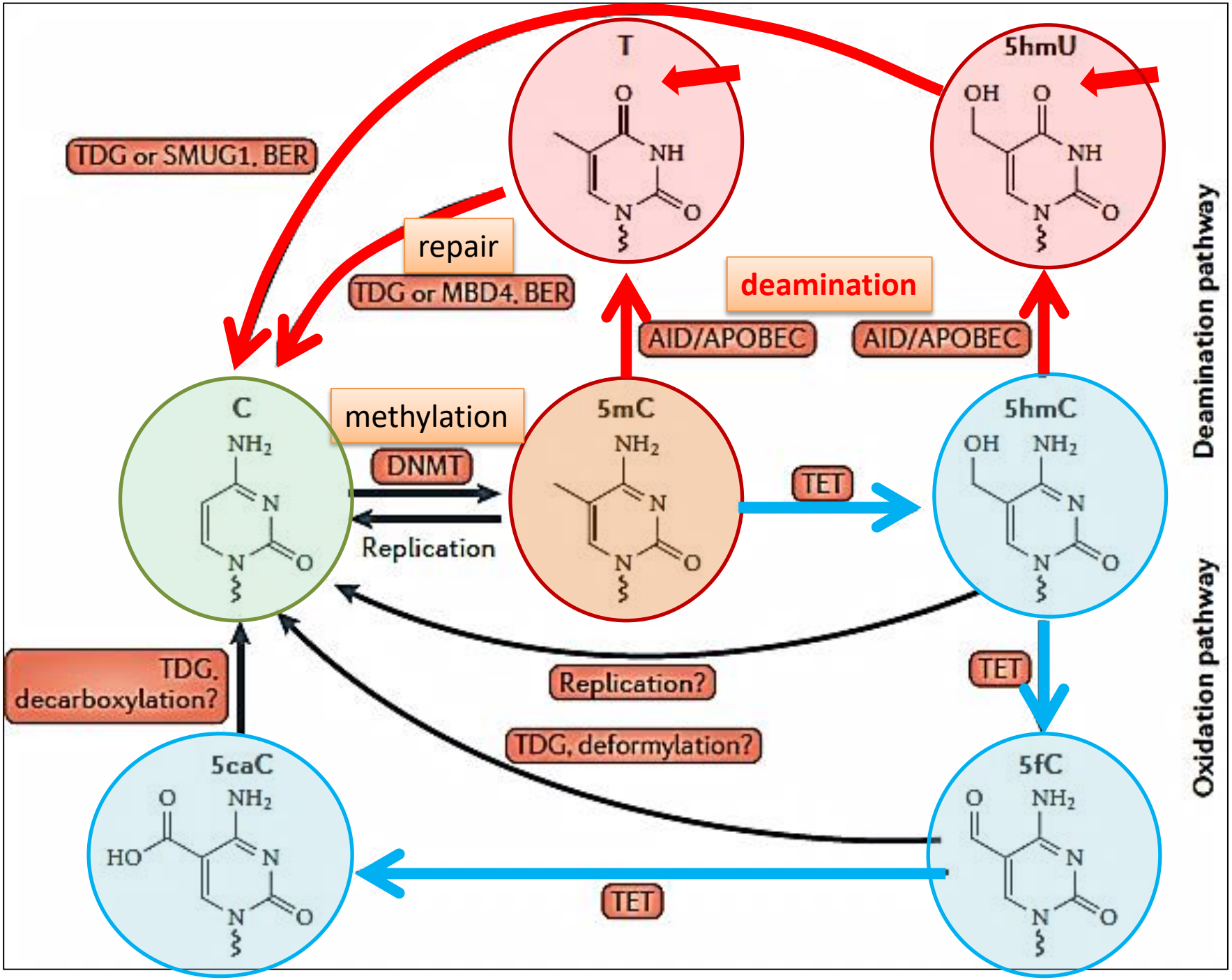


Another epigenetic mark that is very dynamic (contrary to what it was tought years ago) is DNA CpG methylation.

During development or reprogramming, de-novo methylases are found associated with protein complexes containing HDACs, HMTs and other co-repressor complexes.

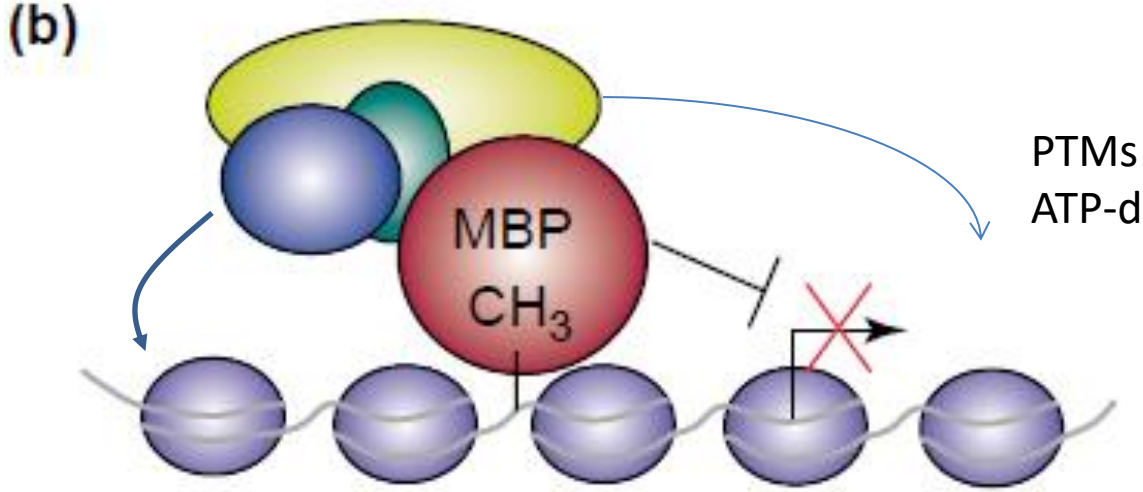
The story of de-methylation is more complex, since as you remember, different mechanisms are in place depending on the context.

Branco et al., Nat Rev Genet. (2012) 13: 7-13.



Methyl-C binding proteins

Co-repressors: what is this ?

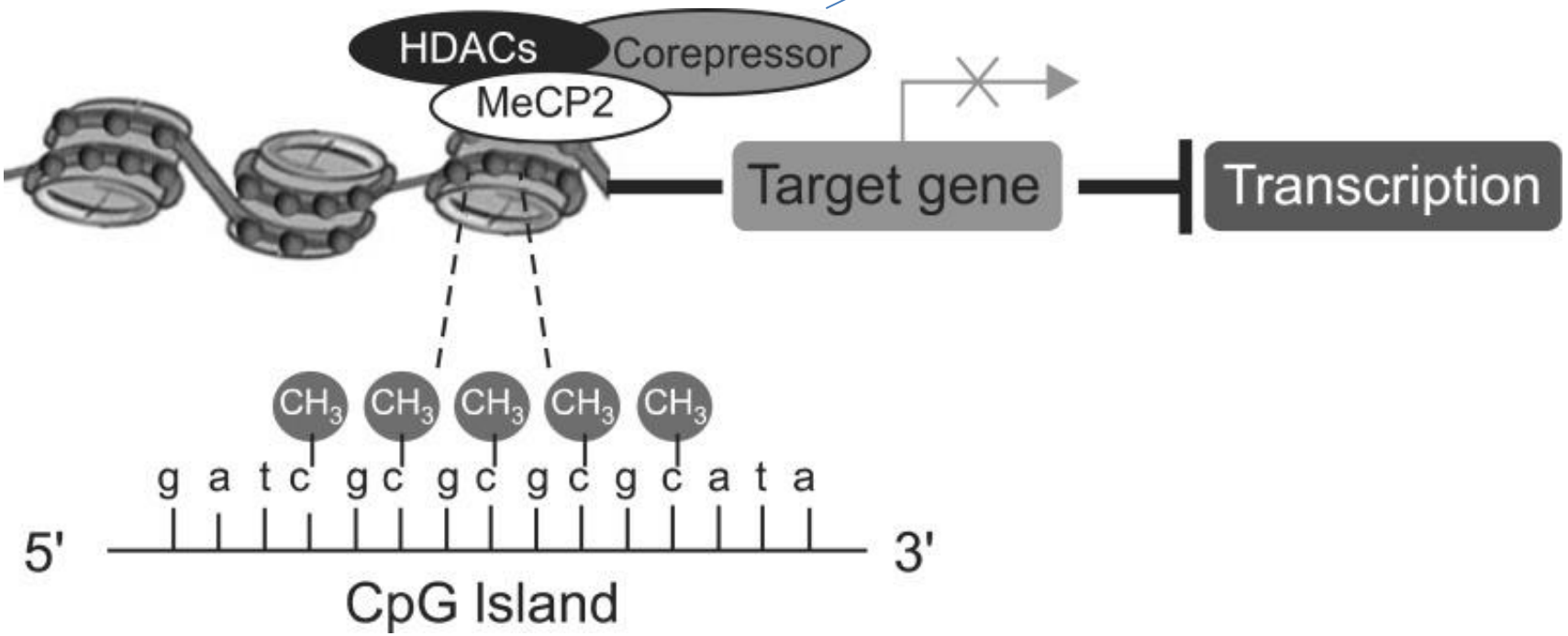


PTMs writers/erasers
ATP-dep. remodeling

Figure 2. Mechanisms of DNA-methylation-mediated repression.
(**b**) Methyl-CpG-binding proteins (MBPs) directly recognize methylated DNA and recruit **co-repressor molecules** to silence transcription and to modify surrounding chromatin

From Klose & Bird, 2006

Corepressors are protein complexes that are present at regulatory regions where they coordinate repressive functions: CpG methylation, HDACs, HMTs and *ATP-dependent chromatin remodeling complexes*



from Seo et al, 2014. doi: 10.9758/cpn.2014.12.2.94

TF interference

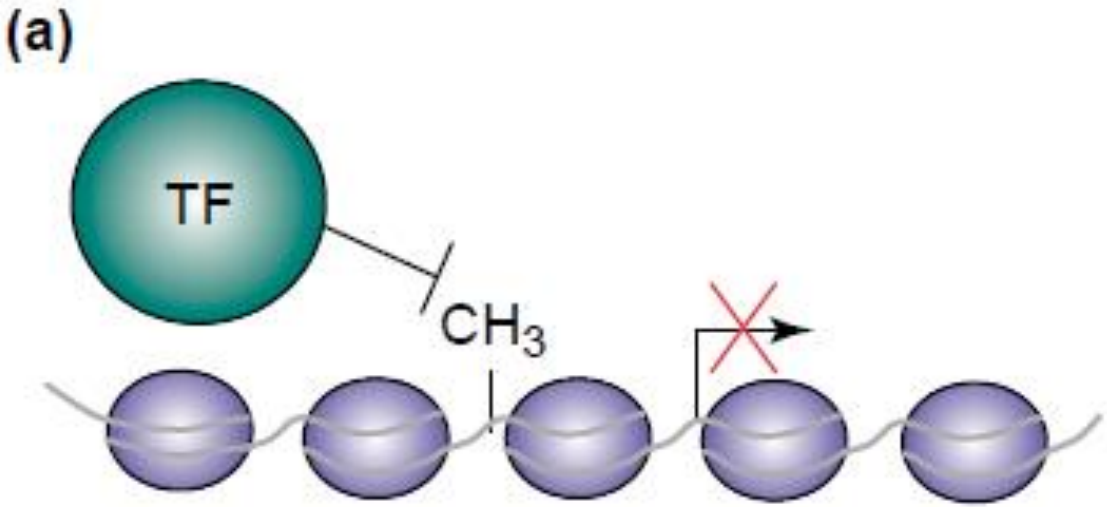
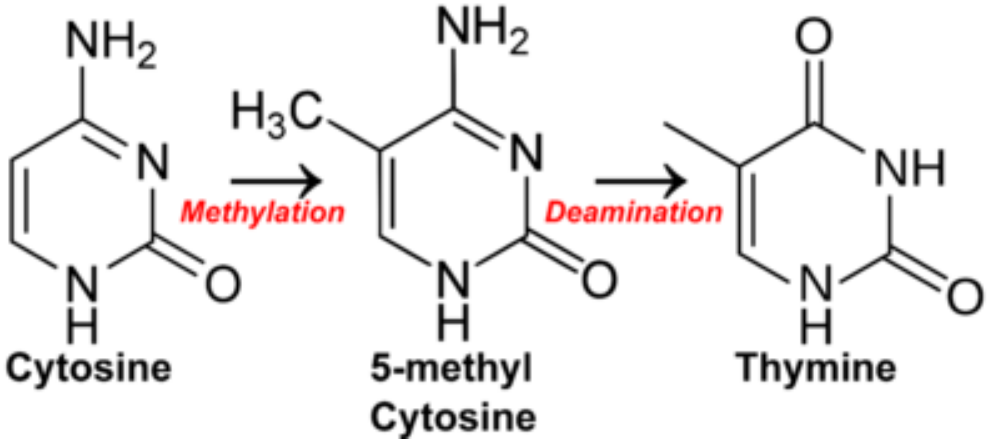


Figure 2. Mechanisms of DNA-methylation-mediated repression.
(**a**) DNA methylation in the cognate DNA-binding sequences of some (*not many, ndr*) transcription factors (TF) can result in inhibition of DNA binding. By blocking activators from binding targets sites, DNA methylation directly inhibits transcriptional activation

From Klose & Bird, 2006

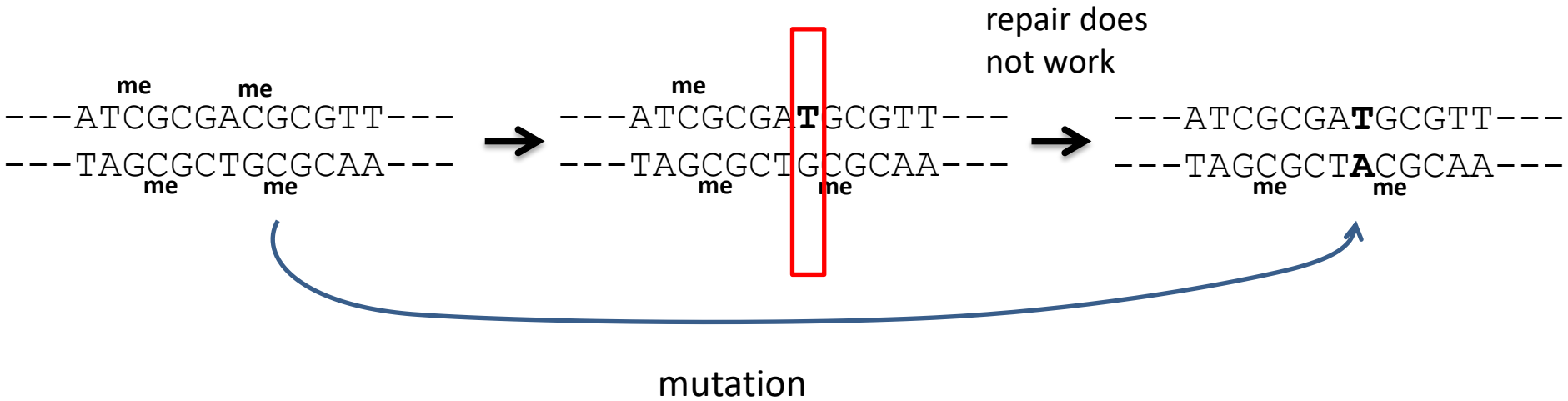
CpG islands and Cytosine Methylation analysis



spontaneous deamination of nucleic acid bases is massive

repair mechanisms are in place to insure sequence maintenance

but ....



CpG-island distribution in Mammalian genomes is limited to functional regions

These were defined as «orphan» but ENCODE has unraveled them being regulatory regions

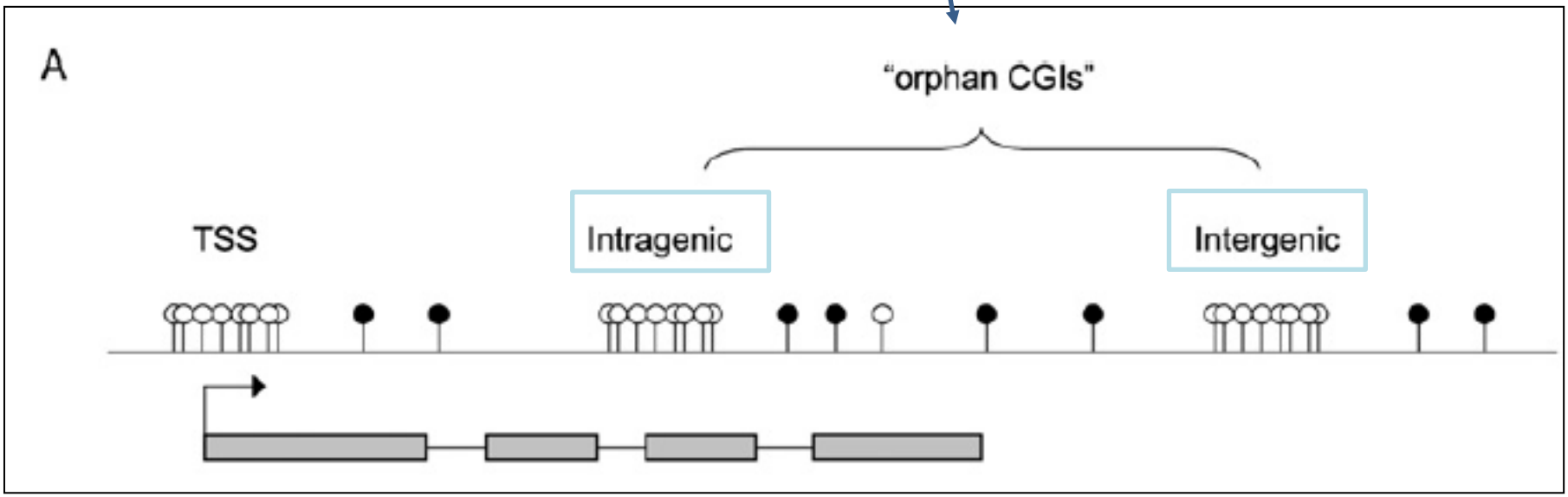


Figure 1. The genomic distribution of CGIs (CpG islands).
(A) CGIs can be located at annotated TSSs, within gene bodies (Intragenic), or between annotated genes (Intergenic).
Intragenic and intergenic CGIs of unknown function are classed as ‘‘orphan’’ CGIs.
(Empty circles) Unmethylated CpG residues. (Filled circles) Methylated CpG residues.

Deaton & Bird, 2011. Genes Dev 25:1010–1022

CpG methylation has been studied genome wide by several laboratories. The first exhaustive study was published by Weber et al. in 2007, using Methyl-DNA immunoprecipitation and analysis on tiling promoter arrays.

Later other groups have used bisulfite-NGS sequencing, and results were confirmed and extended.

Weber et al. first divided the Promoters in three classes, on the basis of the number of CpG in the sequence (CpG/base pair): LCP=low; ICP=intermediate; and HCP=High.

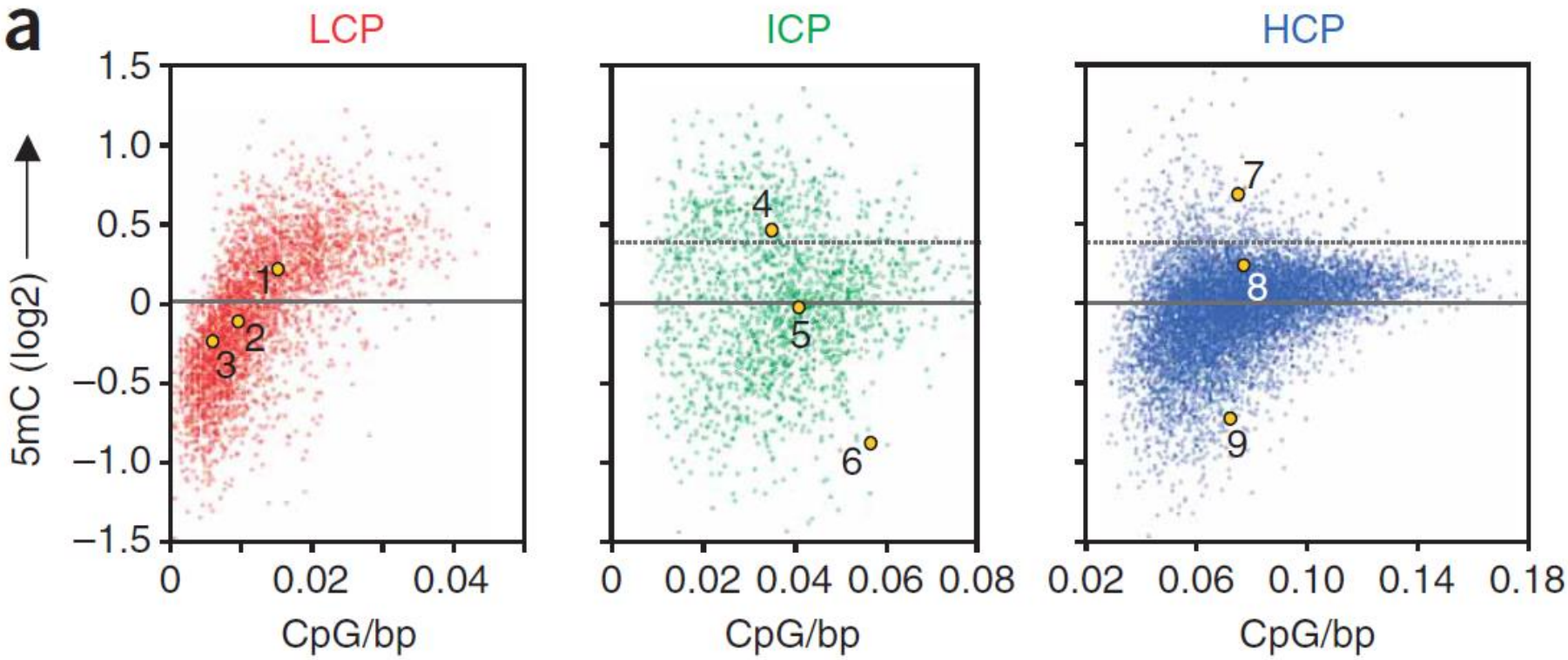
Next, they measured the level of CpG methylation and expressed it in comparison to the density of CpG in each promoter (a window of -700bp to +200bp respect to the TSS).

Finally, they correlated the %-methylation with RNA PolII and H3K4me3 occupancy (by ChIP) as detectors of the activity status.

Enrichment by MeDIP, as an indirect measure of the amount of Cytosine methylation, *versus* the density of CpG in each promoter, in each class.

In LCP, there is a almost linear correlation, i.e. more CpG you have, more meC you get. In HCP, is seems like a «saturation curve» i.e. inreasing most are unmethylated independently on the CpG density.

Numbers indicate single promoters that were studied with bisulfite (next)

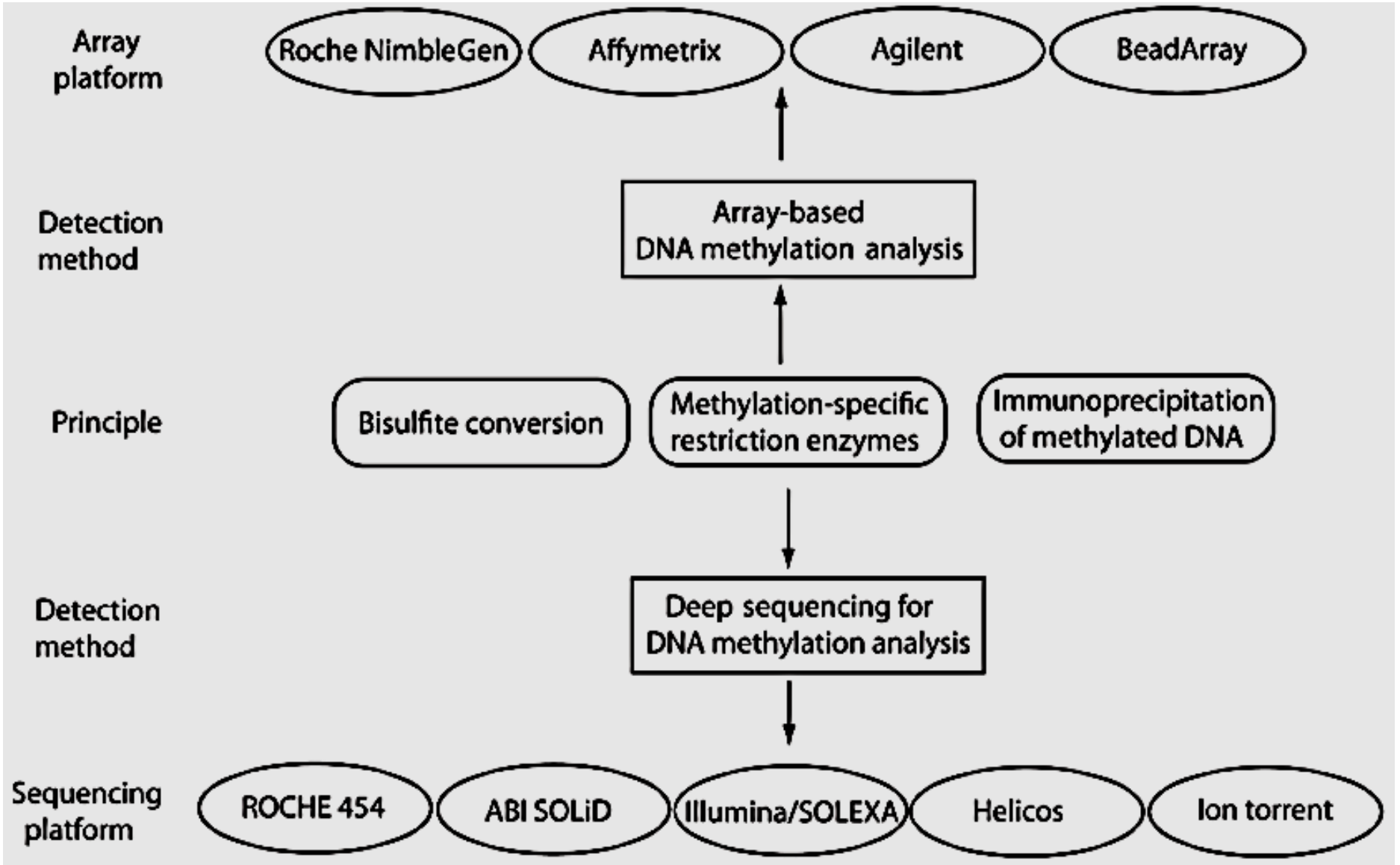


Weber et al, (2007) Nat Genet 39:457.

Importance of DNA methylation for epigenetic inheritance (esp. Mammals)

- CGI (CG-rich islands) are frequent at gene promoters
- there are established molecular links between DNA-methylation machinery and the histone repressive complexes
- CpG-methylation profiles display typical cycles accompanying all developmental stages
- Inhibitors of DNA-methylation induce de-differentiation events
- CGI dysmetabolism is frequently linked to disease (esp. cancer)

methods for genome-wide DNA methylation analysis



Gupta et al., 2010, Biotechniques 49: iii-xi

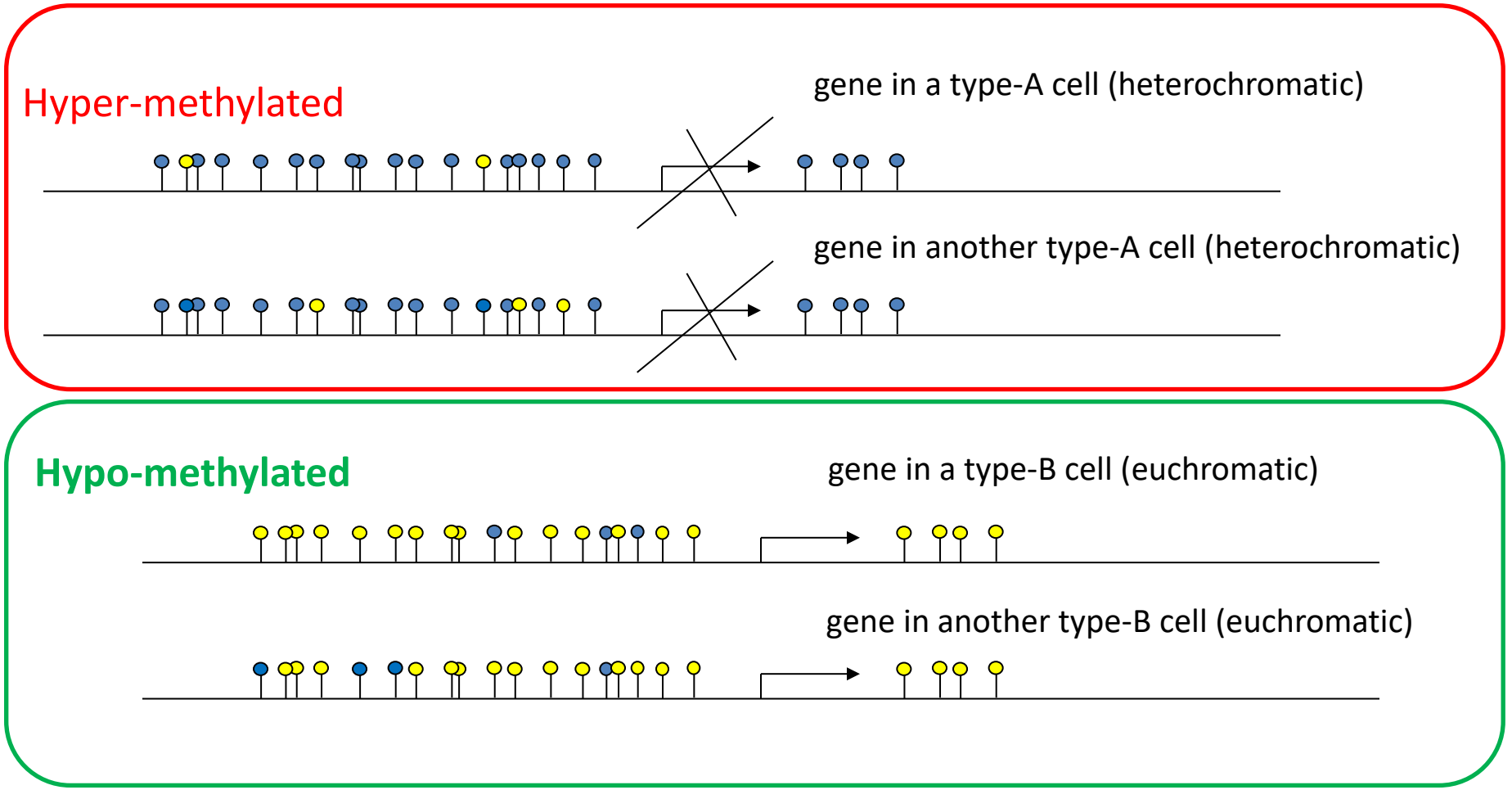
Whole genome methylation by bisulfite modification is **very difficult** !

....remember that CpG methylation is stochastic, every cell has its own profile

Very **redundant** sequencing is needed

Aligning database should contain <u>all possible variations</u>

Warning! – CpG methylation is not site-specific



Since CpG methylation of a given DNA fragment is different from cell to cell, sequencing needs either: **cloning + Sanger** or **NGS**

e.g. (Sanger sequencing) :

1) extract DNA from cells
2) bisulfite treatment
3) PCR the fragment using "side" primers
4) **clone** individual fragment in a plasmid vector
5) sequence a representative number of clones (if Sanger)
   today NGS preferred and equal cost

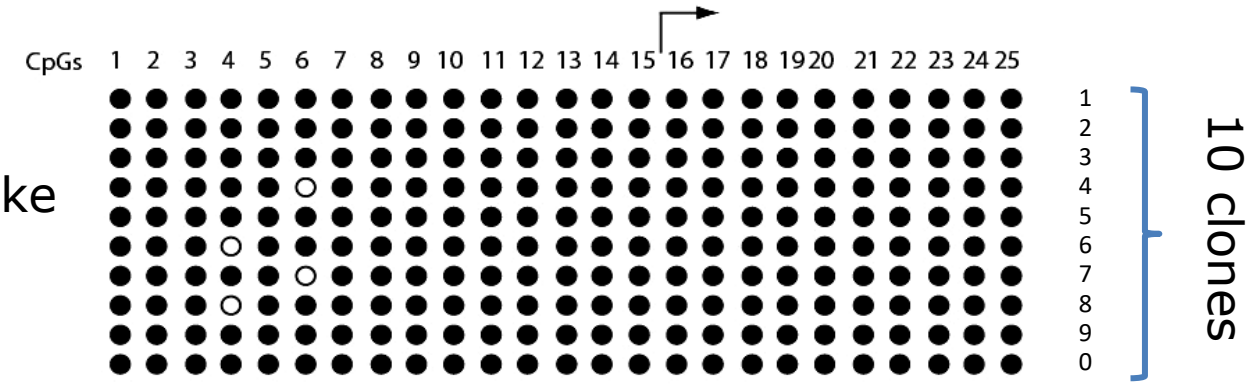
E-Cadherin gene promoter

Example

AGACCCTAGCAACTCCAGGCTAGAGGGTCACCGCGTCTATGCGAGGCCGGGTGGGCGGGCCGTCAGCTCCGCCCTG
GGGAGGGGTCCGCGCTGCTGATTGGCTGTGGCCGGCAGGTGAACCCTCAGCCAATCAGCGGTACGGGGGGCGGTG
CTCCGGGGCTCACCTGGCTGCAGCCACGCACCCCCTCTCAGTGGCGTCGGAACTGCAAAGCACCTGTGAGCTTGCG
GAAGTCAGTTCAGACTCCAGCCCGCTCCAGCCCGGCCCGACCCGACCGCACCCGGCGCCTGCCCTCGCTCGGCGTC
CCCGGCCAGCCATGGGCCCTTGGA

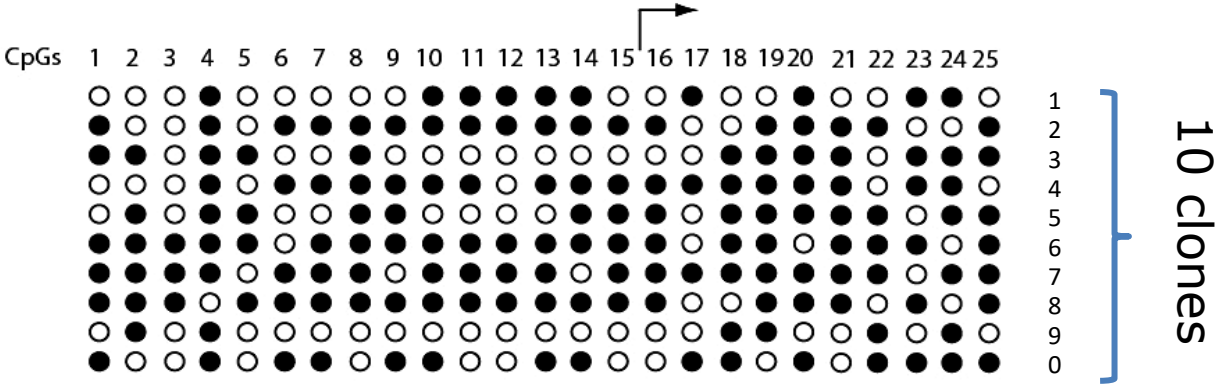
CGI

TSS



Epithelial-like

- ○ methylated
- ● unmethylated



EMT-cells

From: Cardamone et al., 2009, PNAS 106(18):7420-5

Whole genome analysis

Vol 462 | 19 November 2009 | doi:10.1038/nature08514 nature

Human DNA methylomes at base resolution show widespread epigenomic differences

Ryan Lister[1]*, Mattia Pelizzola[1]*, Robert H. Dowen[1], R. David Hawkins[2], Gary Hon[2], Julian Tonti-Filippini[4], Joseph R. Nery[1], Leonard Lee[2], Zhen Ye[2], Que-Minh Ngo[2], Lee Edsall[2], Jessica Antosiewicz-Bourget[5,6], Ron Stewart[5,6], Victor Ruotti[5,6], A. Harvey Millar[4], James A. Thomson[5,6,7,8], Bing Ren[2,3] & Joseph R. Ecker[1]

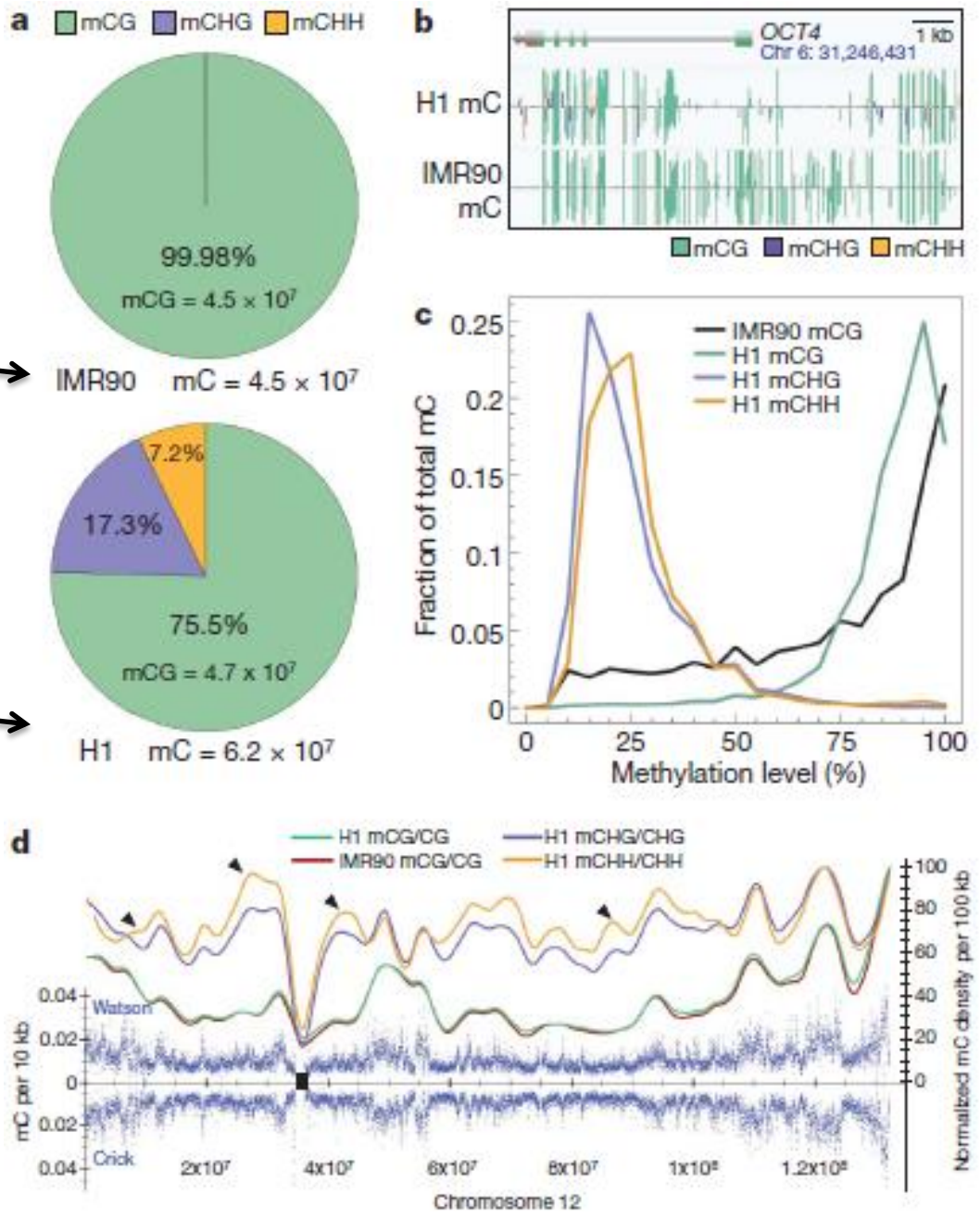
DNA cytosine methylation is a central epigenetic modification that has essential roles in cellular processes including genome regulation, development and disease. Here we present the first genome-wide, single-base-resolution maps of methylated cytosines in a mammalian genome, from both human embryonic stem cells and fetal fibroblasts, along with comparative analysis of messenger RNA and small RNA components of the transcriptome, several histone modifications, and sites of DNA–protein interaction for several key regulatory factors. Widespread differences were identified in the composition and patterning of cytosine methylation between the two genomes. Nearly one-quarter of all methylation identified in embryonic stem cells was in a non-CG context, suggesting that embryonic stem cells may use different methylation mechanisms to affect gene regulation. Methylation in non-CG contexts showed enrichment in gene bodies and depletion in protein binding sites and enhancers. Non-CG methylation disappeared upon induced differentiation of the embryonic stem cells, and was restored in induced pluripotent stem cells. We identified hundreds of differentially methylated regions proximal to genes involved in pluripotency and differentiation, and widespread reduced methylation levels in fibroblasts associated with lower transcriptional activity. These reference epigenomes provide a foundation for future studies exploring this key epigenetic modification in human disease and development.

1 Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

In this method, genomic DNA is treated with sodium bisulphite (BS) to convert cytosine, but not methylcytosine, to uracil, and subsequent high-throughput sequencing. We performed MethylC-Seq for two human cell lines, H1 human embryonic stem cells and IMR90 fetal lung fibroblasts, generating 1.16 and 1.18 billion reads, respectively, that aligned uniquely to the human reference sequence (NCBI build 36/HG18). The total sequence yield was 87.5 and 91.0 gigabases (Gb), with an average read depth of 14.23 and 14.83 per strand for H1 and IMR90, respectively.

IMR90 are fetal lung fibroblasts

H1 are human embryonic **stem** cells



from Lister et al., 2009

IMR90 are fetal lung fibroblasts

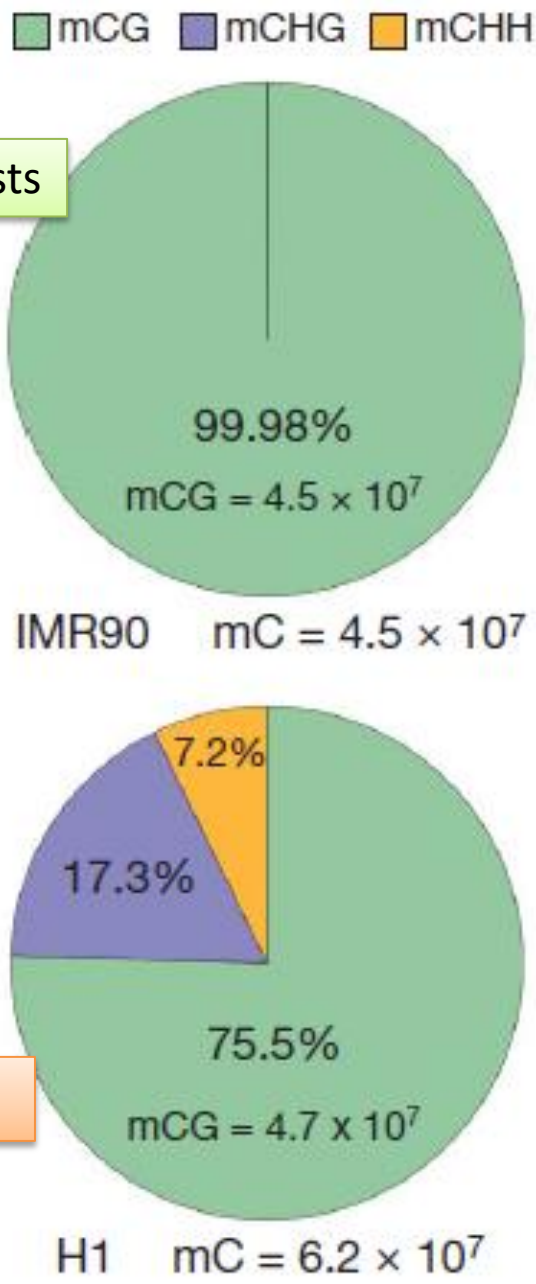
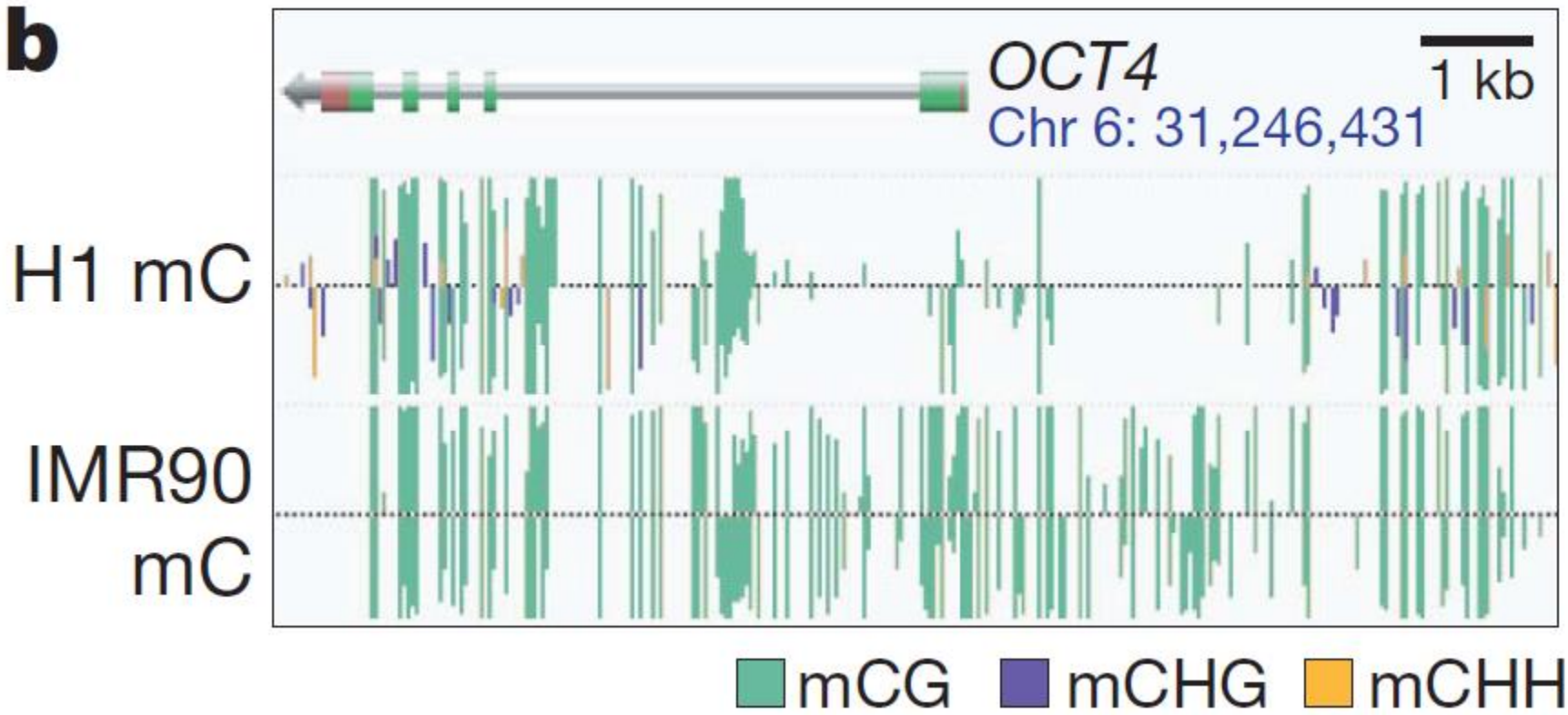


Fig 1 Global trends of human DNA methylomes. The percentage of methyl-cytosines identified for H1 and IMR90 cells in each sequence context

H1 are human embryonic stem cells

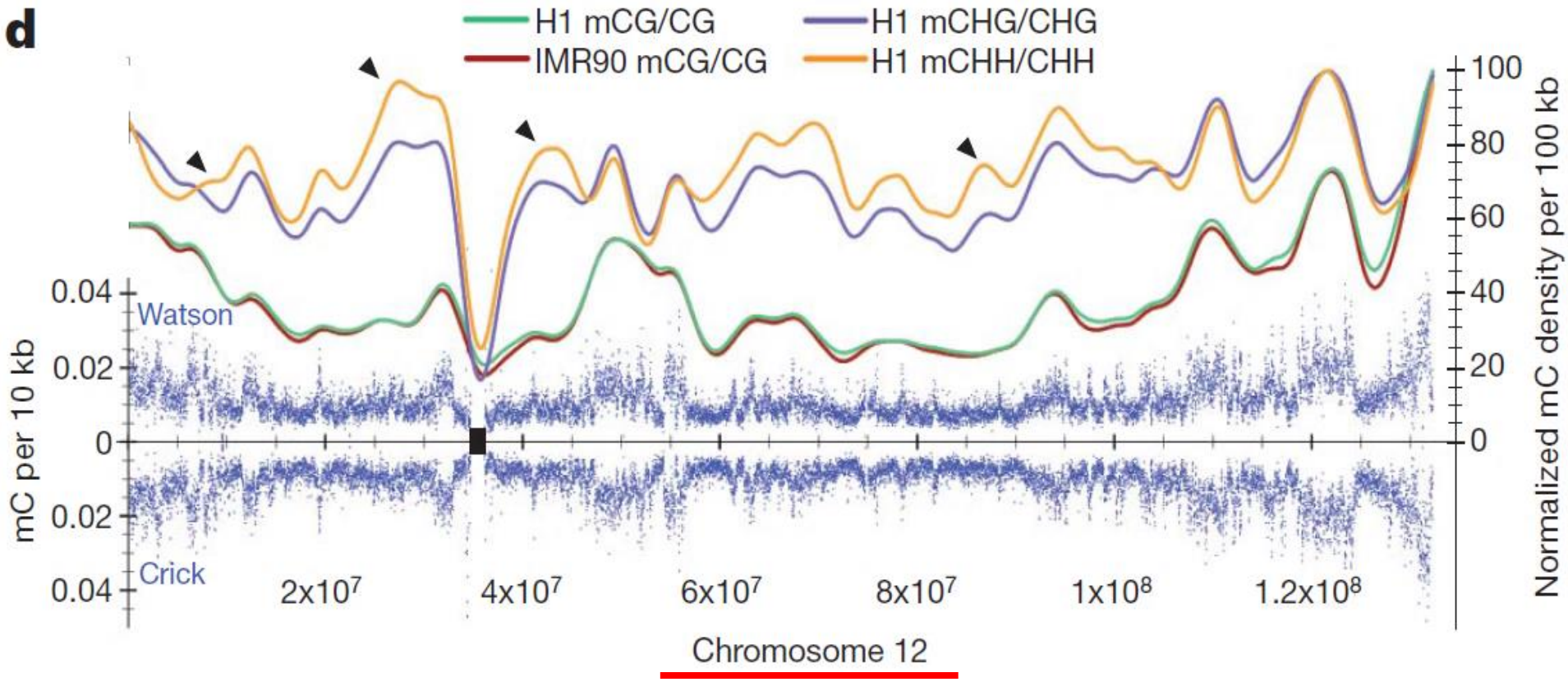
from Lister et al., 2009



b, AnnoJ browser representation of OCT4.

from Lister et al., 2009

from Lister et al., 2009



d, Blue dots indicate methyl-cytosine density in H1 cells in 10-kb windows throughout chromosome 12 (black rectangle, centromere). Smoothed lines represent the methyl-cytosine density in each context in H1 and IMR90 cells. Black triangles indicate various regions of contrasting trends in CG and non-CG methylation. mC, methyl-cytosine.

NATURE|Vol 462|19 November 2009

EPIGENOMICS

Methylation matters

Dirk Schübeler

Genome-wide maps of methylated cytosine bases at single-base-pair resolution in human cells reveal distinct differences between cell types. These maps provide a starting point to decode the function of this enigmatic mark.

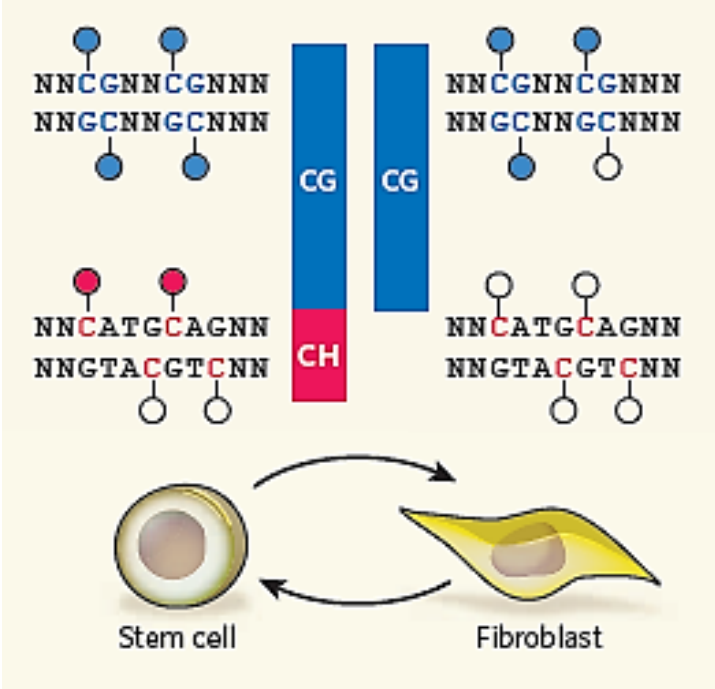


Figure 1 | DNA methylation patterns differ between stem cells and differentiated cells[1]. In stem cells, regions of DNA with CpG methylation (blue) are mostly uniformly methylated, whereas this modification is more heterogeneous in fibroblasts. Non-CpG methylation (red), which occurs primarily at CA nucleotides, is detected only in stem cells, yet is asymmetric and more scarce and patchy than CpG methylation. If fibroblasts are converted to induced pluripotent stem cells they regain non-CpG methylation. Filled circles, methylated cytosines; unfilled circles, unmethylated cytosines. H stands for A, C or T; N stands for any nucleotide.

Single-base-resolution maps of DNA methylation for two human cell lines

Single-base DNA methylomes of the flowering plant *Arabidopsis thaliana* were previously achieved using MethylC-Seq[15] or BS-Seq[16]. In this method, genomic DNA is treated with sodium bisulphite (BS) to convert cytosine, but not methylcytosine, to uracil, and subsequent high-throughput sequencing. We performed MethylC-Seq for two human cell lines, H1 human embryonic stem cells[17] and IMR90 fetal lung fibroblasts[18], generating 1.16 and 1.18 billion reads, respectively, that aligned uniquely to the human reference sequence (NCBI build 36/HG18). The total sequence yield was 87.5 and 91.0 gigabases (Gb), with an average read depth of 14.2× and 14.8× per strand for H1 and IMR90, respectively (Supplementary Fig. 1a). In each cell type, over 86% of both strands of the 3.08 Gb human reference sequence are covered by at least one sequence read (Supplementary Fig. 1b), accounting for 94% of the cytosines in the genome.

from Lister et al., 2009

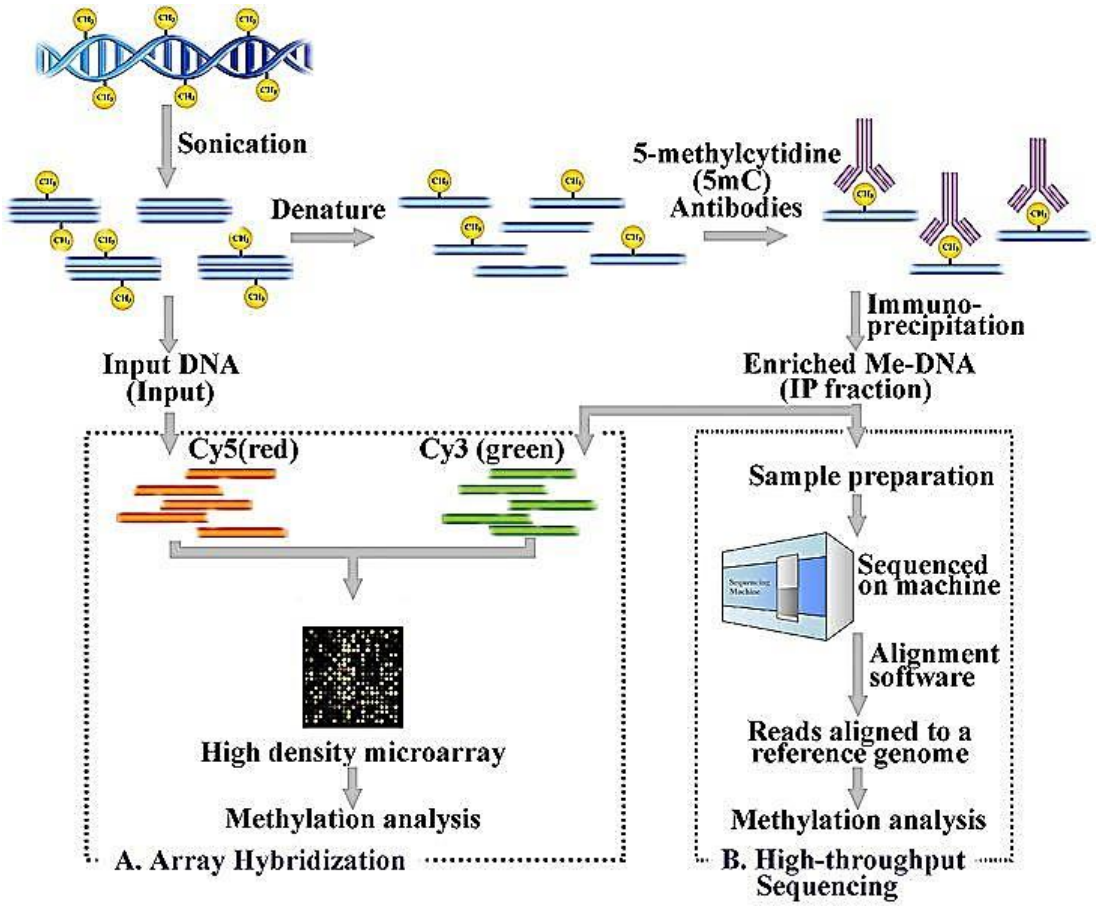
RRBS = Reduced Representation Bisulfite Sequencing

1

DNA cut with CCGG-specific R.E. s will enrich a 200-300 bp fraction of CGI DNA. This fraction is used to produce library → NGS

2

Or affinity-purified (MeDIP) methylated DNA fragments analyzed:



RRBS workflow